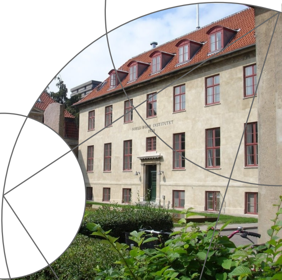
Cleaning and Training

Data-driven Photon Particle Identification In the ATLAS Experiment

Benjamin Henckel, Troels C. Petersen



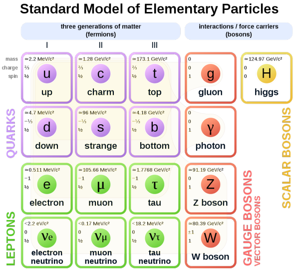
Outline

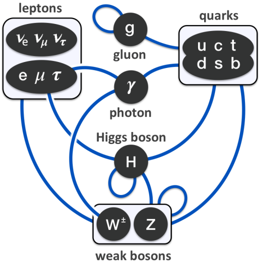
- **Theory**
- **The ATLAS Experiment**
- **Machine Learning Methods**
- **Datasets**
- **Models**
- **MC Results**
- **Training In Data**
- **Label Confusion**
- **Decorrelation Framework**
- **Data Results**
- **Evaluation on $H \rightarrow \gamma\gamma$**
- **Summary & Outlook**

Aim of the thesis

e/γ Particle Identification at the ATLAS Experiment with focus on data-driven methods.

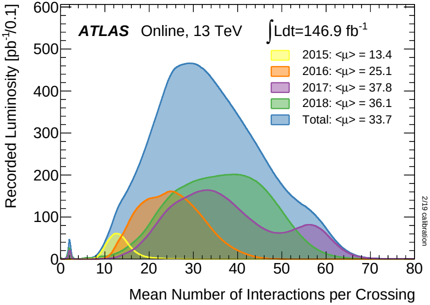
Theory



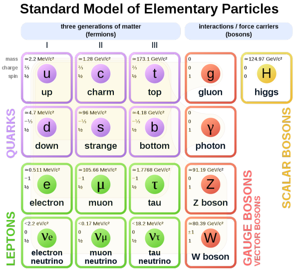


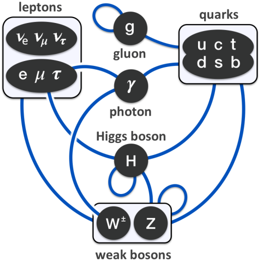
Pileup is a measure of how much is happening in your detector.

Luminosity is a direct measure of your ability to probe interesting physics



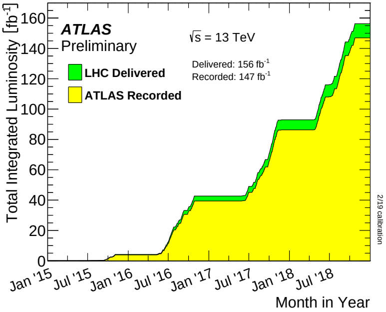
Theory



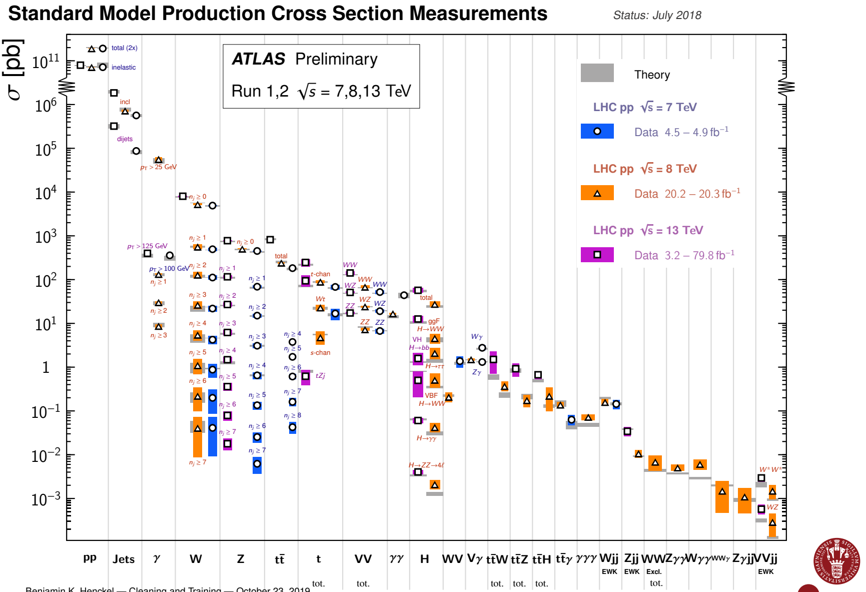


Pileup is a measure of how much is happening in your detector.

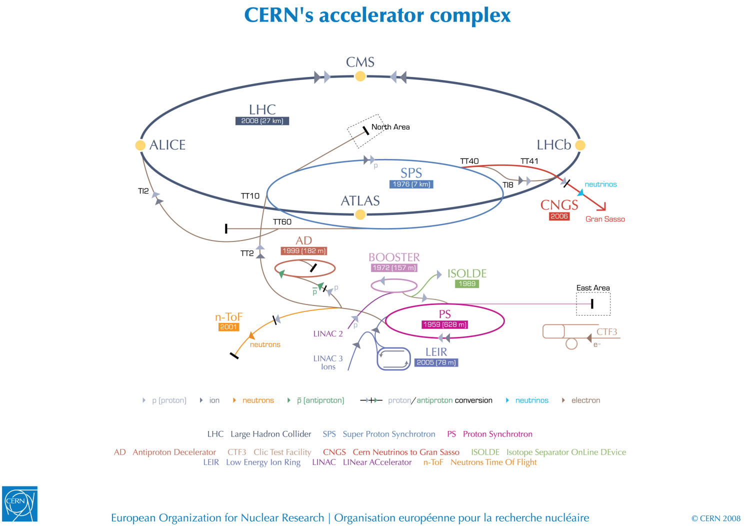
Luminosity is a direct measure of your ability to probe interesting physics



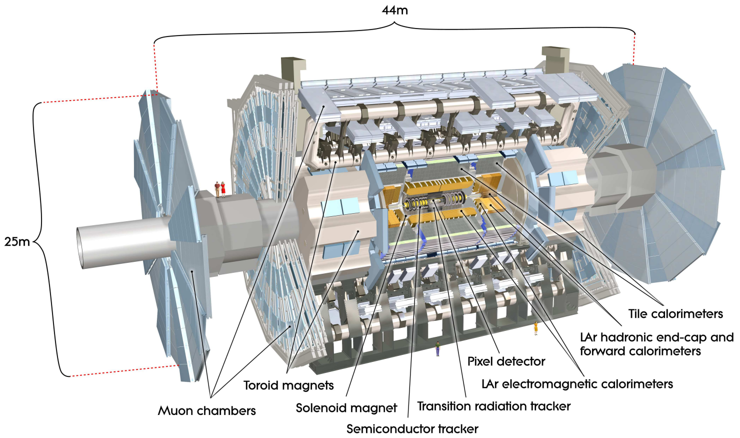
Theory


Standard Model Production Cross Section Measurements
Status: July 2018
ATLAS Preliminary
Run 1,2 √s = 7,8,13 TeV
σ [pb]
Theory
LHC pp √s = 7 TeV
Data 4.5 – 4.9 fb⁻¹
LHC pp √s = 8 TeV
Data 20.2 – 20.3 fb⁻¹
LHC pp √s = 13 TeV
Data 3.2 – 79.8 fb⁻¹
pp
Jets
γ
W
Z
tt̄
t
VV
γγ
H
WV
Vγ
tt̄W
tt̄Z
tt̄H
tt̄γ
γγγ
Wjj EWK
Zjj EWK
WWExcl.
Zγγ
Wγγ
WWγ
Zγjj
VVjj EWK
tot.

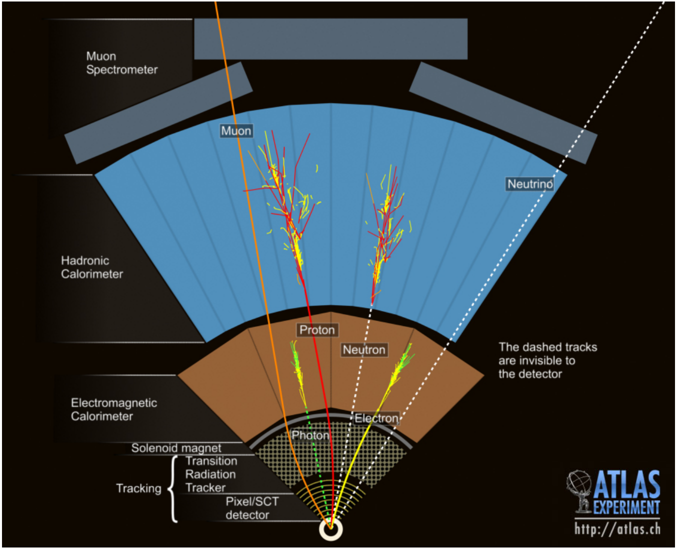
The ATLAS Experiment



The ATLAS Experiment



The ATLAS Experiment



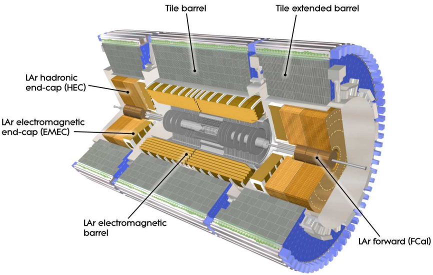
The ATLAS Experiment

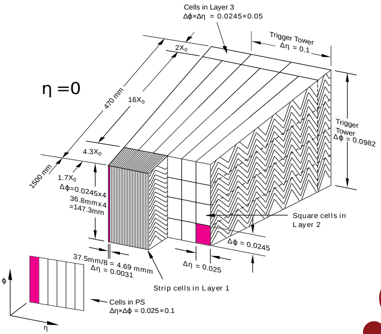
The calorimeter measures the energy of the particles by stopping them.

It is seperated in electromagnetic (ECAL) and hadronic (HCAL) subdetectors.

The entirety of the ECAL uses a LAr as the active material and have accordion shaped lead absorbers.

The HCAL is a sampling calorimeter. It uses LAr in the end-cap and scintillating material in the rest.





The ATLAS Experiment

Photon Identification

Loose

1 dimensional cuts on 4 variables

Optimized in bins of $|\eta|$

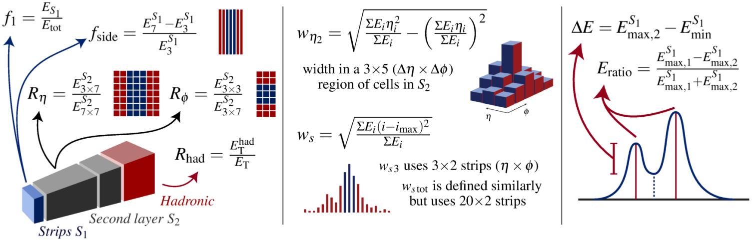
Designed to select all signal ($\approx 99\%$), while reducing background by a factor 1000

Tight

1 dimensional cuts on 10 variables

Optimized in bins of $|\eta|$ and E_T

Designed for analyses.



Machine Learning Methods

Different types of Machine Learning models have been trained. Tree based models: **Random Forest** (RF) and **Boosted Decision Trees** (BDTs) and feed-forward **Neural Networks** (NN).

The trees based models are implemented using the LightGBM framework provided by Microsoft.

The Neural Networks are implemented using Keras with TensorFlow backend.

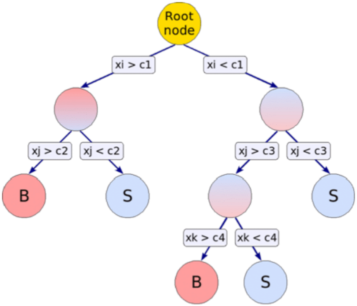
Many other frameworks have been used for crucial parts in the model development.

Optimization is done using the *gp_minimize* and *rf_minimize* packages from Sci-kit Optimize.

Feature importance has been estimated using SHapely Additive exPlanations (SHAP).

Reweighting has been done using the GBReweighter package from *hep_ml*.

Machine Learning Methods



Random Forest

Creates independent decision trees

Using **bagging** and **subsampling**

Majority vote makes a strong classifier

Boosted Decision Trees

Creates dependent decision trees

Using **Boosting** and **gradient descent**

Continued training will lead to over training!

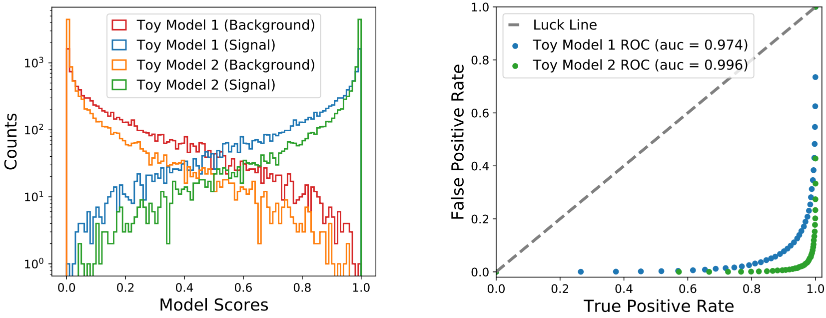
Machine Learning Methods: Evaluation

Four possible outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

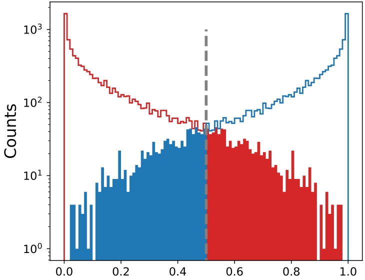
$$TPR = \frac{TP}{TP + FN} \quad \text{and} \quad FPR = \frac{FP}{FP + TN}$$

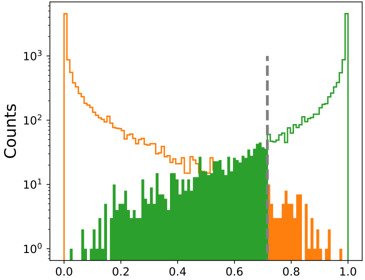
The **ROC** curve is a visualization of TPR and FPR.

AUC reduces the ROC curve to one number.



ML Methods





Improvement factors are defined:

$$Imp_{FPR} = \frac{FPR_{m1}}{FPR_{m2}} - 1 = \frac{FP_{m1}}{FP_{m2}} - 1 \quad and \quad Imp_{TPR} = \frac{TPR_{m1}}{TPR_{m2}} - 1 = \frac{TP_{m1}}{TP_{m2}} - 1$$

Datasets

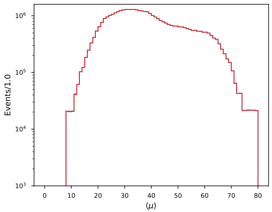
MC Photon Dataset

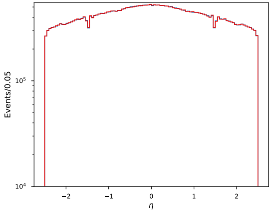
High statistics (104,976,458 datapoints) dataset taking advantage of the truth matching available in MC.

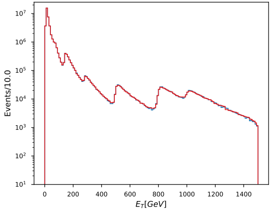
The only selection applied $p_T < 4.5 GeV$

Covers a massive energy range, but has many potential flaws.

Label	Total [%]	Composition [%] and Description	
Signal	7.91	100.0	Isolated Photons
Background	92.09	70.31	Background Photons
		19.30	Non truthmatched Objects
		8.23	Hadrons
		1.36	Electrons
		0.73	Non Isolated Photons







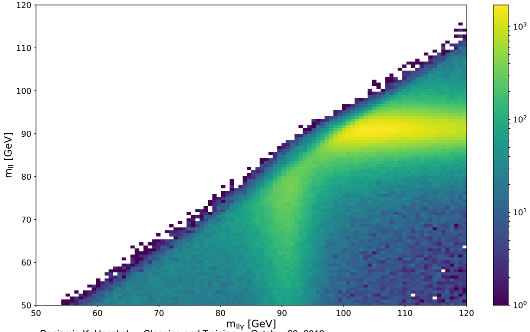
Datasets

DATA Photon Dataset

A Tag, Tag, and Probe (TTP) framework intended to select $Z \to ll\gamma$ in DATA.

Uses all of data17 (9,576,799 probes selected)

The 2D histogram of the invariant mass of the tag pair and of the tag pair plus the probe reveals significant background in output.



TagElectronPairs
- Both have pT > 26 GeV
- Both have track
- Both from first primary vertex
- Both pass Medium ID
- Both pass Loose Isolation
- Triggers

TagMuonPairs
- Both have pT > 26 GeV
- Both have track in ID
- Both from first primary vertex
- Both pass medium quality
- Both pass Loose isolation
- Triggers

probePhotons
- Has pT > 9.5 GeV
- Pass Object Quality (cluster)

llGamma Candidates
- DeltaR(Tag, Probe) > 0.4 (0.2) for electron (muon) tags

N(Candidate) = 1

Selected Events

Models

Models will be trained on an **increasing** number of variables.

The variables are based on the ones directly and indirectly used by the current methods.

A few **new**, carefully chosen, variables have been added in the extra variable set.

Name	Nvars	Vars
Showershape (pdf)	11	Rhad1 Rhad Reta weta1 weta2 Rphi Reta fracs1 DeltaE Eratio f1
Conversion (conv)	15	ConvertionType ConversionRadius VertexConvEtOverPt VertexConvPtRatio
Binning (bin)	18	η E_T $\langle\mu\rangle$
Extra (ext)	22	*maxEcell_time* *maxEcell_energy* core57cellsEnergyCorrection r33over37allcalo

The namingscheme of the models follow this outline:

$$Modelname : ''MODELTYPE''\,(''DATATYPE'' , ''varlist'')$$

Models

The isolation models are trained on the following 9 variables:
*topoetcone*20, *topoetcone*30, *topoetcone*40,
*ptvarcone*20, *ptvarcone*30, *ptvarcone*40,
E_T, *NvtxReco*, *averageInteractionsPerCrossing*

For training and evaluation of all models the datasets have been split into training, validation and evaluation sets

Training set is shown to the models during training

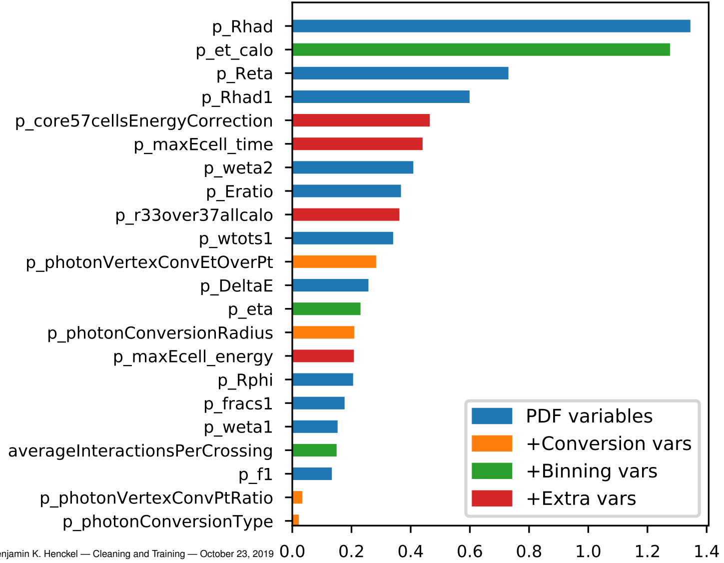
Validation set is used for early stopping

Evaluation set is only used for making final plots

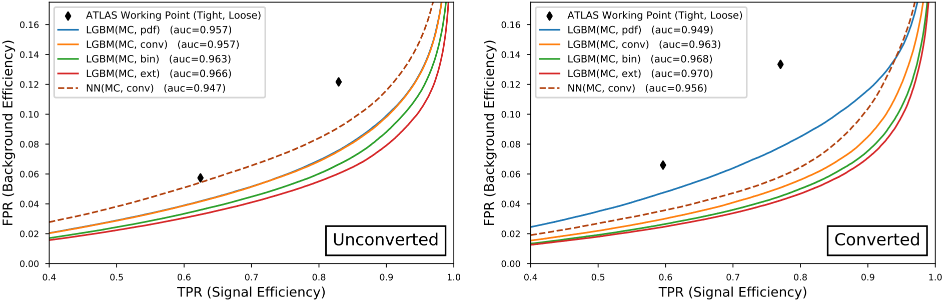
The training is done **inclusively** in E_T, η and conversions

MC Trained Model Evaluation

MC Results

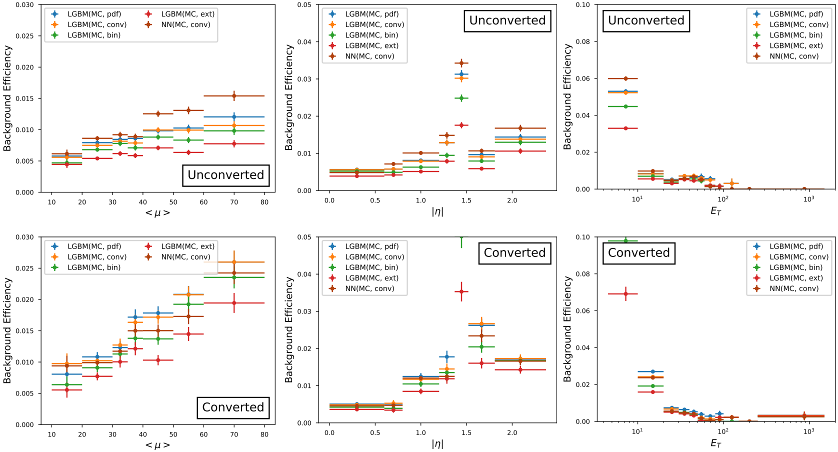

p_Rhad
p_et_calo
p_Reta
p_Rhad1
p_core57cellsEnergyCorrection
p_maxEcell_time
p_weta2
p_Eratio
p_r33over37allcalo
p_wtots1
p_photonVertexConvEtOverPt
p_DeltaE
p_eta
p_photonConversionRadius
p_maxEcell_energy
p_Rphi
p_fracs1
p_weta1
averageInteractionsPerCrossing
p_f1
p_photonVertexConvPtRatio
p_photonConversionType
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
PDF variables
+Conversion vars
+Binning vars
+Extra vars

MC Results



The results in MC will be evaluated seperately for the Unconverted and Converted channels. The ROC curves reveal a good overall performance across all models, in both channels.

MC Results



The background efficiency as a function of $\langle\mu\rangle$, $|\eta|$, and E_T, at a fixed signal efficiency of 90% against **hadronic** background.

Training in DATA

Biggest advantage is training directly on real variable distributions

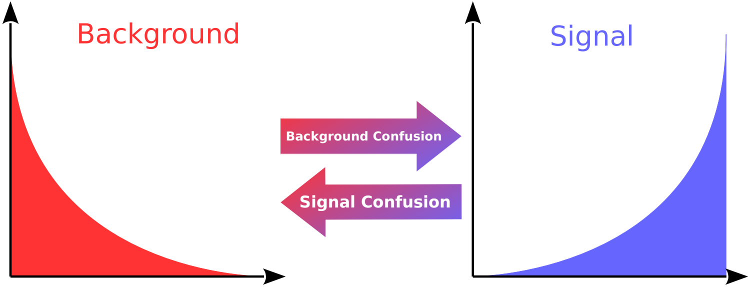
Biggest disadvantage is lack of true labels

This leaves two major questions to be answered:

How does our models behave when given imperfect labels?

How do we aquire labels in DATA?

Label Confusion

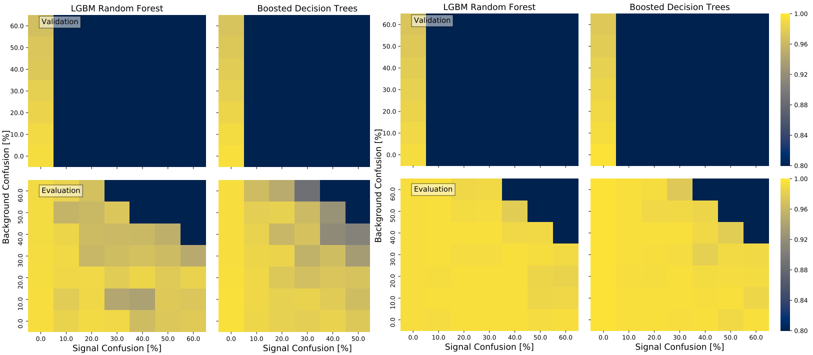


Label Confusion describes the presence of wrongly assigned signal and background labels in the dataset. This is inescapable in DATA.

Label Confusion

(a) Low Statistics Training

(b) High Statitics Training

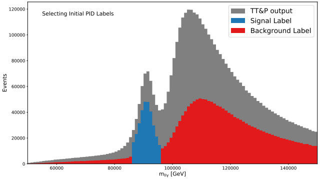


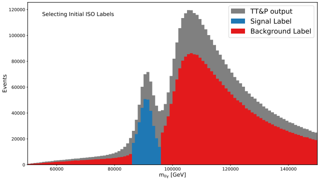
Evolution of **AUC** as a function of increasing confusion, for validation (top) and evaluation (bottom) sets.

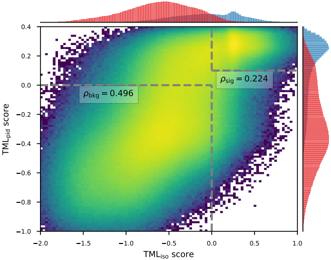
Decorrelation Framework

Initial labels are chosen using the following cuts:

Task	Signal Selection	Background Selection
Identification	$\lvert m_{\ell\ell} - m_Z \rvert < 10GeV, \quad \frac{etcone40}{E_T} < 0.06$	$\lvert m_{\ell\ell} - m_Z \rvert > 10GeV, \quad \frac{etcone40}{E_T} > 0.15$
Isolation	$\lvert m_{\ell\ell} - m_Z \rvert < 10GeV, \quad Loose$	$\lvert m_{\ell\ell} - m_Z \rvert > 10GeV, \quad !Loose$







The logit transformation is the following:

$$TML = -log(1/ML - 1)/k$$

Decorrelation Framework

A linear decorrelation usually applies a rotation, which in turn changes both values. The intend is to only change the isolation variable. This is achieved by defining the following variables:

$$x = \frac{pid}{\sigma_{pid}} \quad and \quad y = \frac{iso}{\sigma_{iso}}$$

Then the following is true:

$$\sigma^2(x) = \sigma^2(y) \quad and \quad Cov(x, y) = \rho_{xy}$$

We can the define a new y variable that is linearly uncorrelated from x:

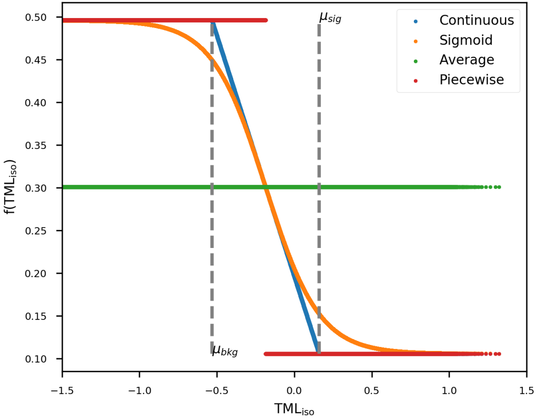
$$y' = y - \rho \times x$$

Inserting x and y, we find the following recipe for an isolation variable that is linearly uncorrelated from particle identification:

$$iso' = iso - \rho_{iso,pid} \times \frac{\sigma_{iso}}{\sigma_{pid}} \times pid \quad ^1$$

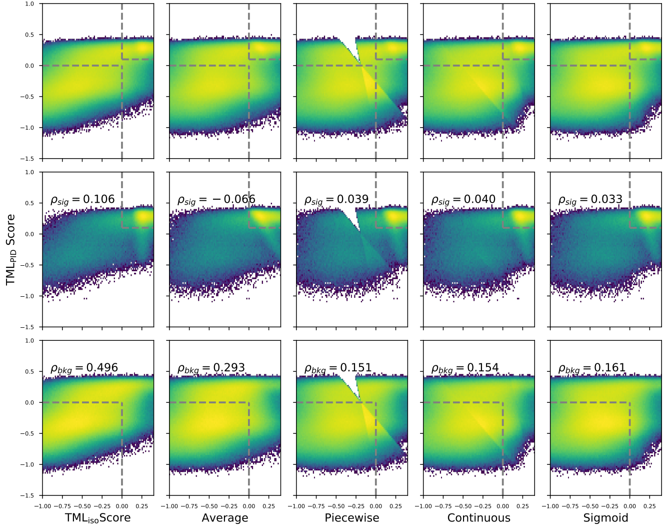
[1] Mistakenly used variance in thesis

Decorrelation Framework



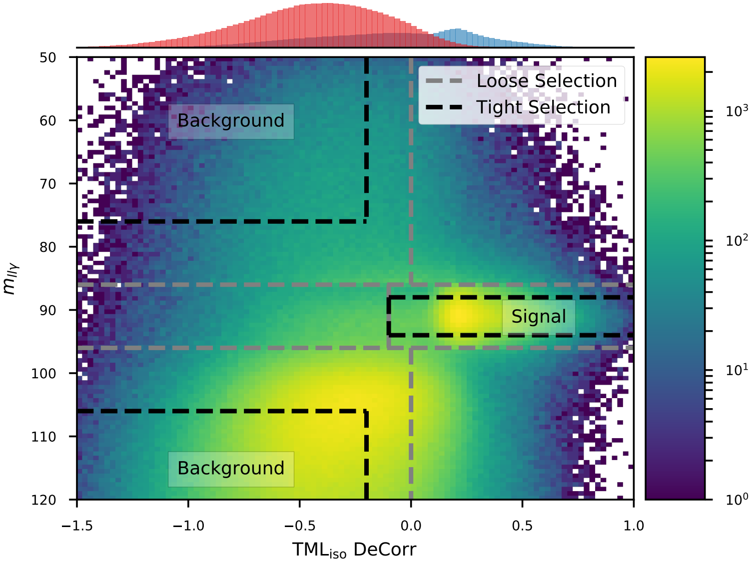
A visualization of the decorrelation methods. $\rho_{sig,bkg}$, σ_{iso} and σ_{pid} are calculated following these recipes.

Decorrelation Framework



The resulting decorrelation, when applying the different methods.

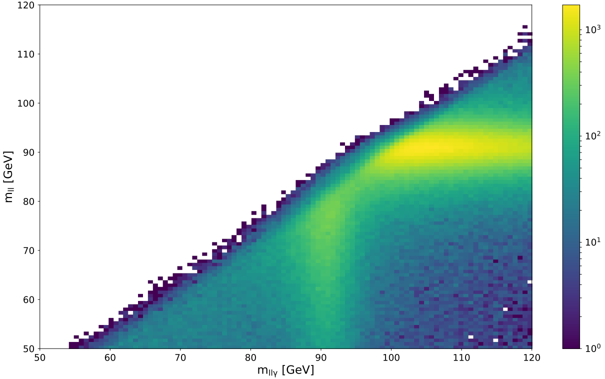
Decorrelation Framework



Results in DATA

The background efficiency was matched in a background region ($m_{\ell\ell\gamma} > 110 GeV$)

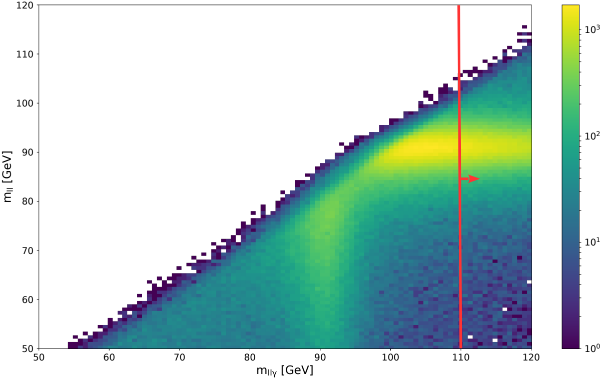
For evaluation $m_{\ell\ell} < 82 GeV$ cut was applied. Improvement calculated from all events left after cut.



Results in DATA

The background efficiency was matched in a background region ($m_{\ell\ell\gamma} > 110 GeV$)

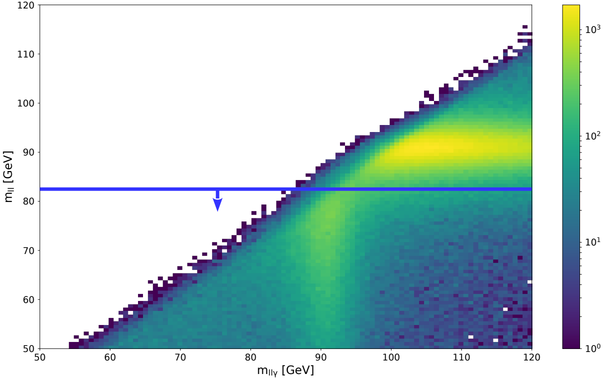
For evaluation $m_{\ell\ell} < 82 GeV$ cut was applied. Improvement calculated from all events left after cut.



Results in DATA

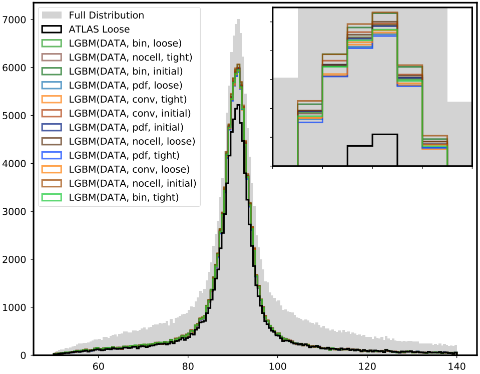
The background efficiency was matched in a background region ($m_{\ell\ell\gamma} > 110GeV$)

For evaluation $m_{\ell\ell} < 82GeV$ cut was applied. Improvement calculated from all events left after cut.

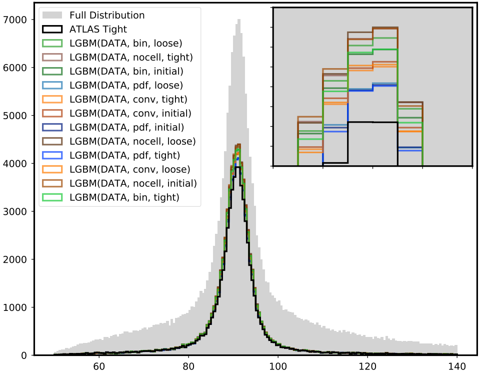


Results in DATA

Loose



Tight



The **DATA** trained models evaluated on the $Z \to \ell\ell\gamma$ peak, with background efficiency matched to the Loose (left) and Tight (right) working points.

Results in DATA

Improvement [%] (ATLAS Loose)

DATA Models	pdf	conv	bin	nocell
initial	**12.10 ± 0.05**	12.60 ± 0.05	13.94 ± 0.06	**14.33 ± 0.06**
loose	**11.40 ± 0.05**	11.89 ± 0.05	13.12 ± 0.06	**13.39 ± 0.06**
tight	**11.12 ± 0.05**	11.60 ± 0.05	12.72 ± 0.05	**12.94 ± 0.06**
MC Models				
Truth	12.59 ± 0.05	12.92 ± 0.06	13.78 ± 0.06	14.08 ± 0.06

Improvement factor of all models evaluated on the $Z \to \ell\ell\gamma$ peak, with background efficiency matched to the Loose (top) and Tight (bottom) working points. The first column indicate the label selection the model was trained on.

Results in DATA

Improvement [%] (ATLAS Loose)

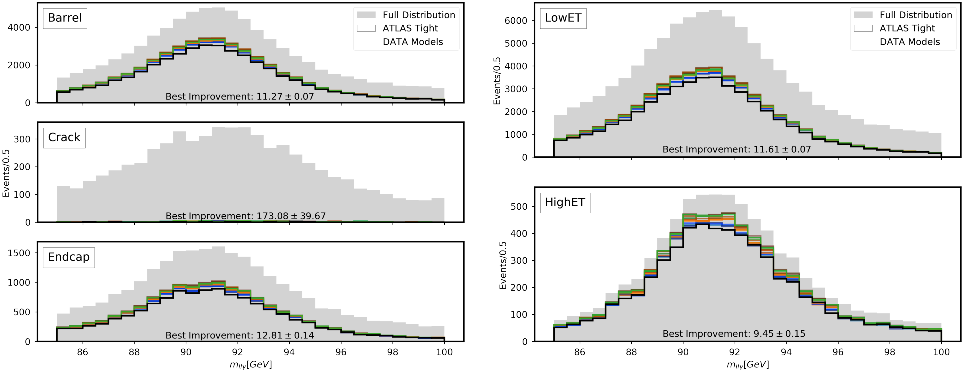
DATA Models	pdf	conv	bin	nocell
initial	**12.10 ± 0.05**	12.60 ± 0.05	13.94 ± 0.06	**14.33 ± 0.06**
loose	**11.40 ± 0.05**	11.89 ± 0.05	13.12 ± 0.06	**13.39 ± 0.06**
tight	**11.12 ± 0.05**	11.60 ± 0.05	12.72 ± 0.05	**12.94 ± 0.06**
MC Models				
Truth	12.59 ± 0.05	12.92 ± 0.06	13.78 ± 0.06	14.08 ± 0.06

Improvement [%] (ATLAS Tight)

DATA Models	pdf	conv	bin	nocell
initial	**4.61 ± 0.03**	7.45 ± 0.04	9.06 ± 0.05	**10.87 ± 0.06**
loose	**4.77 ± 0.03**	7.17 ± 0.04	9.92 ± 0.05	**11.11 ± 0.06**
tight	**4.19 ± 0.02**	6.97 ± 0.04	9.05 ± 0.05	**10.78 ± 0.06**
MC Models				
Truth	1.42 ± 0.01	5.74 ± 0.03	6.47 ± 0.04	7.96 ± 0.04

Improvement factor of all models evaluated on the $Z \rightarrow \ell\ell\gamma$ peak, with background efficiency matched to the Loose (top) and Tight (bottom) working points. The first column indicate the label selection the model was trained on.

Results in DATA



Split DATA $Z \to \ell\ell\gamma$ evaluation of DATA models in the barrel ($|\eta| < 1.37$), crack ($1.37 < |\eta| < 1.52$), and end-cap ($1.52 < |\eta| < 2.37$), and at low and high E_T. Compared to the Tight (right) working point.

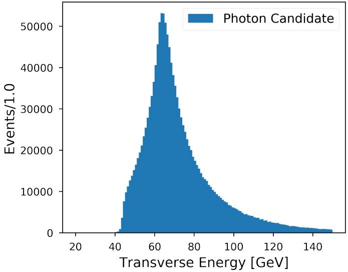
Evaluation on $H \rightarrow \gamma\gamma$

The $H \rightarrow \gamma\gamma$ MC dataset is selected by a simple selection, that does not include background.

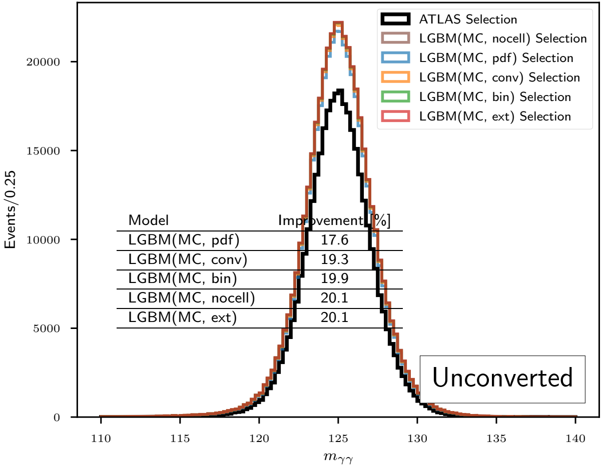
The background efficiencies of the models are matched to the **Tight** working point in the MC photon dataset in the energy range $50 GeV < E_T < 80 GeV$.

This is a favourable comparison for the models, but the best one available.

Candidate Selection
Overlap Removal
$E_T > 25 GeV$
$E_T / m_{\gamma\gamma} > 0.35(0.25)$
Veto Crack
Loose Isolation

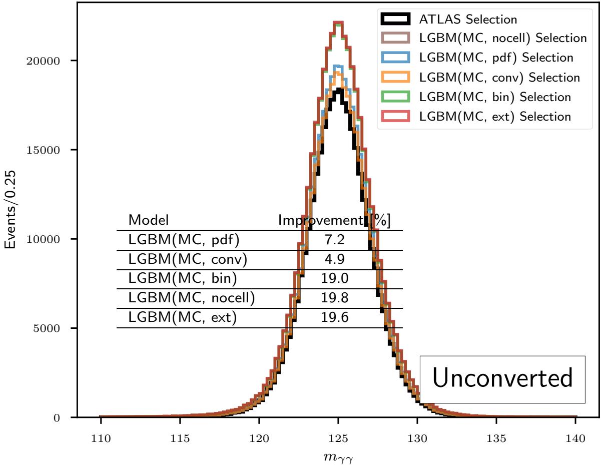


Evaluation on $H \rightarrow \gamma\gamma$



The unconverted channel shows incredibly promising improvements.

Evaluation on $H \rightarrow \gamma\gamma$



Model	Improvement [%]
LGBM(MC, pdf)	7.2
LGBM(MC, conv)	4.9
LGBM(MC, bin)	19.0
LGBM(MC, nocell)	19.8
LGBM(MC, ext)	19.6

To increase confidence in the improvements, the background efficiency has been matched to the Tight working point against **hadronic** background only.

Summary & Outlook

Several MC- and DATA-driven models have been trained and evaluated

A method for selecting high purity samples in DATA has been developed, using an ML isolation model

Evaluation in DATA $Z \to \ell\ell\gamma$, MC $H \to \gamma\gamma$ shows promising improvements.

Summary & Outlook

Several MC- and DATA-driven models have been trained and evaluated

A method for selecting high purity samples in DATA has been developed, using an ML isolation model

Evaluation in DATA $Z \rightarrow \ell\ell\gamma$, MC $H \rightarrow \gamma\gamma$ shows promising improvements.

Implement models in Athena $\rightarrow$ Test them through proper channels

Use data from whole of Run2 for training of DATA models.

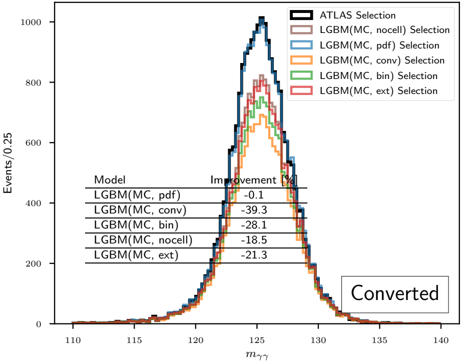
Test on full $H \rightarrow \gamma\gamma$ analysis or similar high profile channel (Most importantly, with background).

Expand onto more complex models.

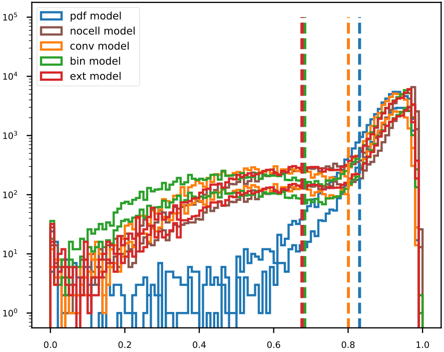
Appendix

$H \rightarrow \gamma\gamma$ Bug in Converted Channel

Evaluation on $H \to \gamma\gamma$



Model	Improvement [%]
LGBM(MC, pdf)	-0.1
LGBM(MC, conv)	-39.3
LGBM(MC, bin)	-28.1
LGBM(MC, nocell)	-18.5
LGBM(MC, ext)	-21.3



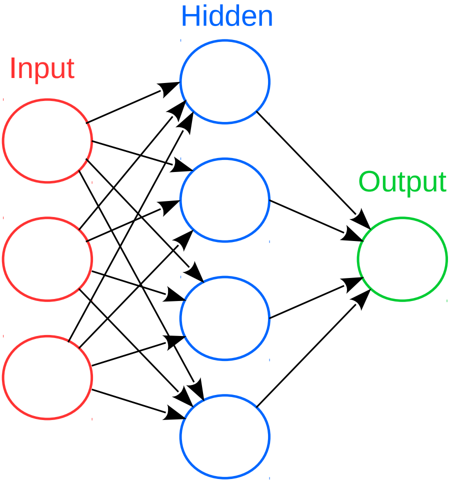
Unfortunately a bug caused the calculation of the conversion variables to fail. This causes all the models that include these variables to colapse. It is hard to conclude anything from the converted channel

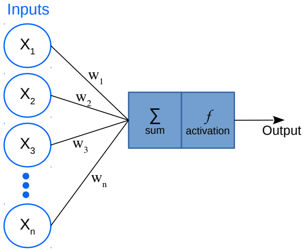
Neural Networks

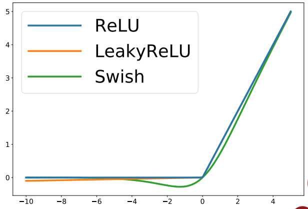
Machine Learning Methods

Neural Network

Universal Function Approximators.







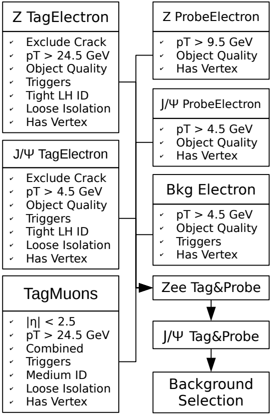
Electron PID Results

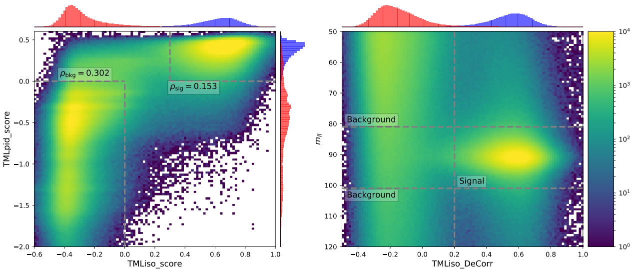
Electron PID results

The Electron dataset is a signal $Z \to ee$ and accompanying background selection run on a large subset of data from 2017.

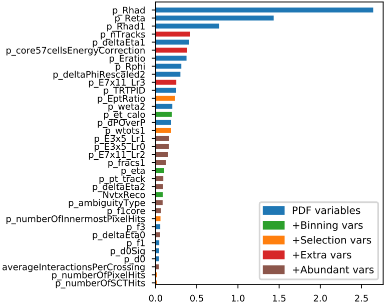
It was used as a test for the decorrelation framework, however, models trained perform well.

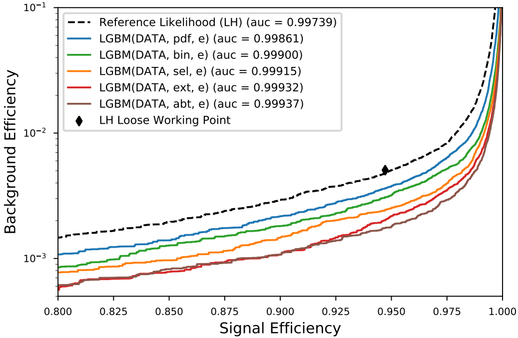
Training in DATA yielded 3 new variables, which did not perform well on MC.





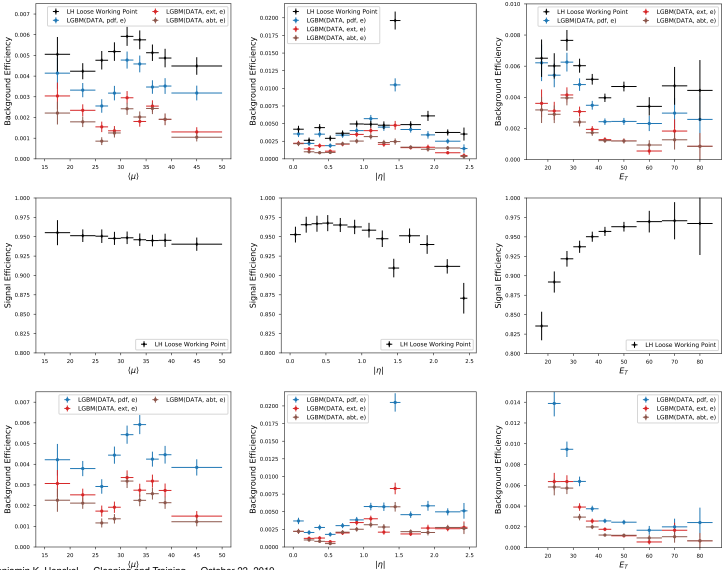
Electron PID results





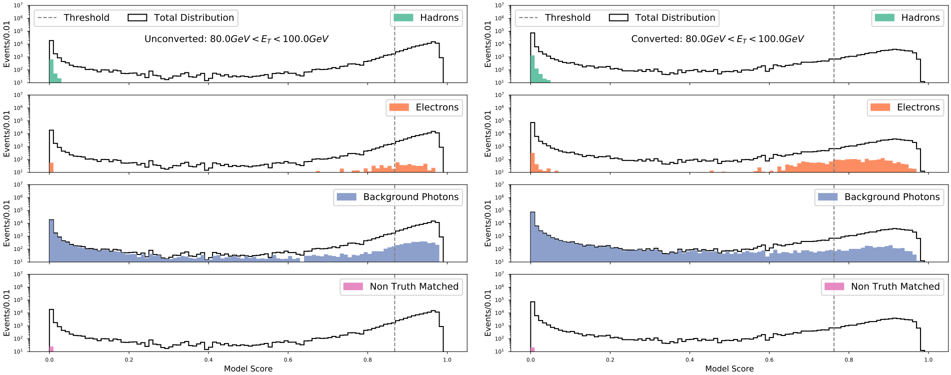
The great performance of the 3 variables in the extra variable set is unique to training in DATA. All models show promising overall performance.

Electron PID results



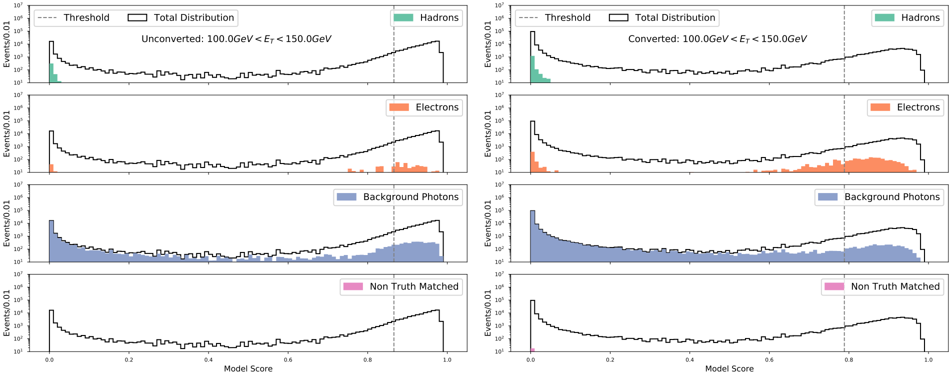
Background Composition in MC Photon Dataset

Background Composition In MC Photon Dataset



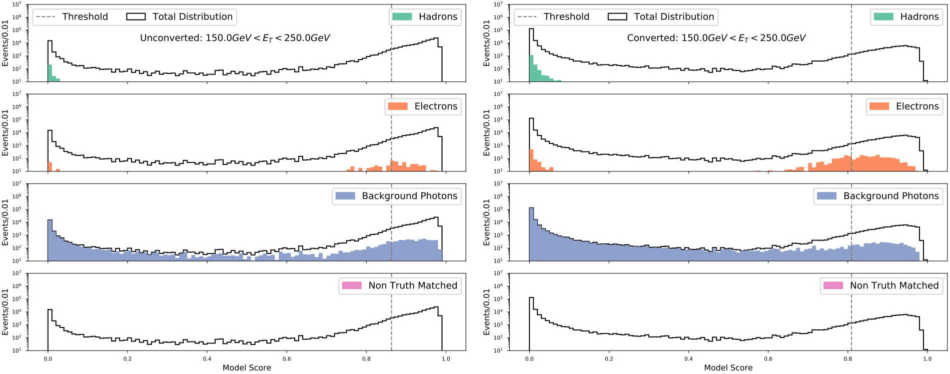
The problematic behaviour in the unconverted channel is caused by background photons, with increasing energy, they become increasingly signal-like. This effect is barely present in the converted channel.

Background Composition In MC Photon Dataset



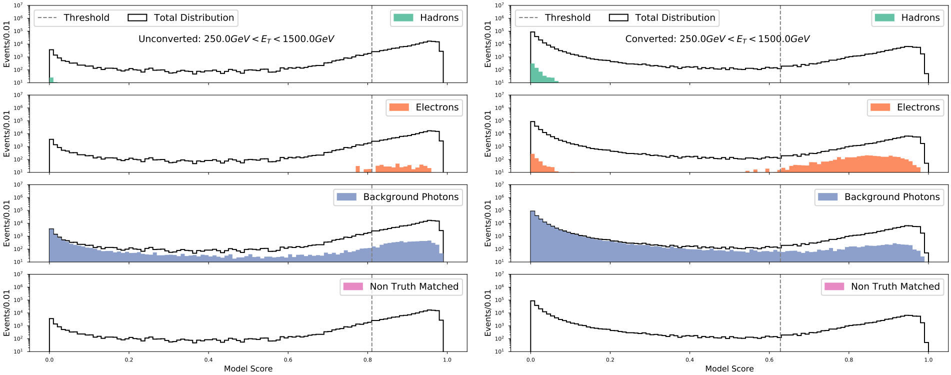
The problematic behaviour in the unconverted channel is caused by background photons, with increasing energy, they become increasingly signal-like. This effect is barely present in the converted channel.

Background Composition In MC Photon Dataset



The problematic behaviour in the unconverted channel is caused by background photons, with increasing energy, they become increasingly signal-like. This effect is barely present in the converted channel.

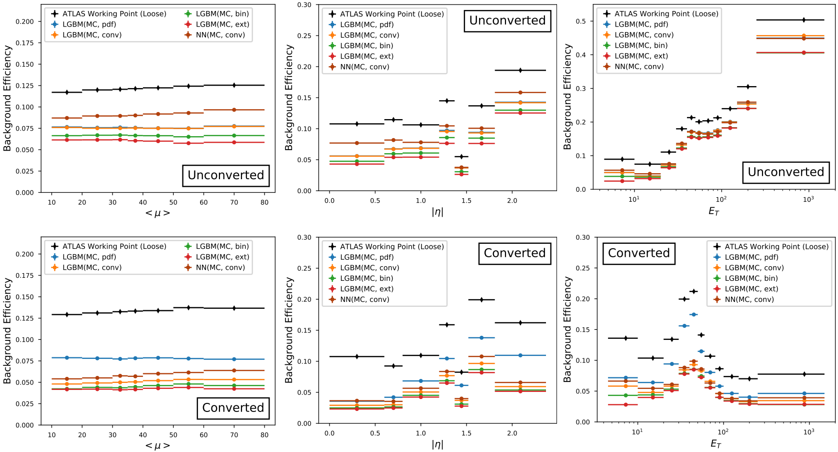
Background Composition In MC Photon Dataset



The problematic behaviour in the unconverted channel is caused by background photons, with increasing energy, they become increasingly signal-like. This effect is barely present in the converted channel.

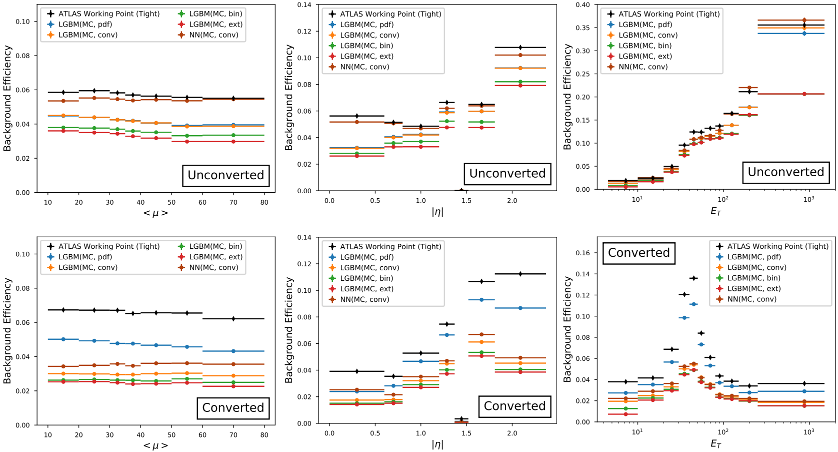
MC Photon Results

MC Results



The background efficiency as a function of $\langle\mu\rangle$, $|\eta|$, and E_T, at the same signal efficiency as the **Loose** working point.

MC Results

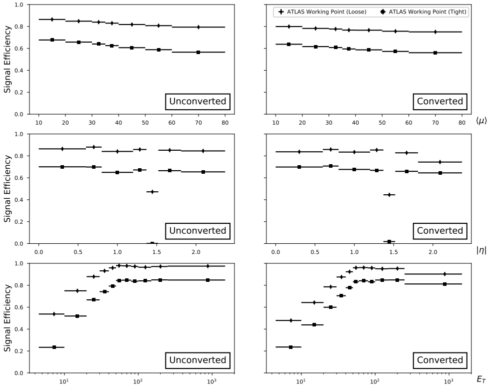


The background efficiency as a function of $\langle \mu \rangle$, $|\eta|$, and E_T, at the same signal efficiency as the **Tight** working point.

MC Results

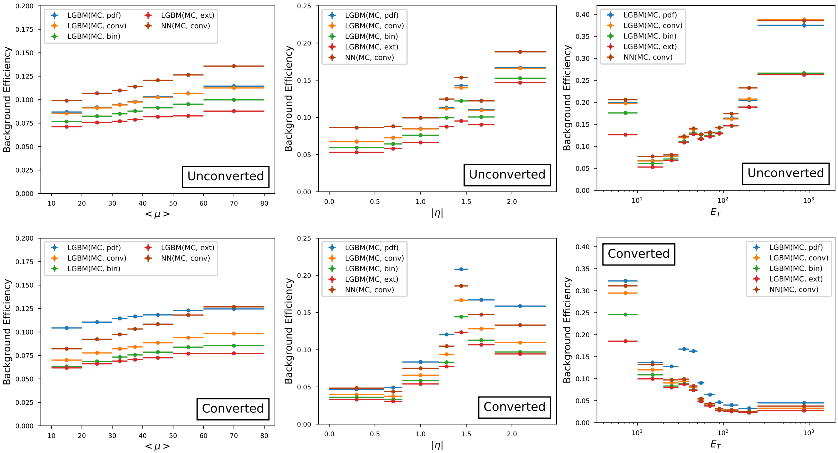
The odd behaviour of the working points (mimiced by the models) seem to be due to background composition.

Supported by the signal efficiency of the working points.



The models should be more robust to this, as they are optimized against this problematic background composition. To gauge this the signal efficiency of the models are fixed at 90%.

MC Results



The background efficiency as a function of $\langle\mu\rangle$, $|\eta|$, and E_T, at a fixed signal efficiency of 90%.

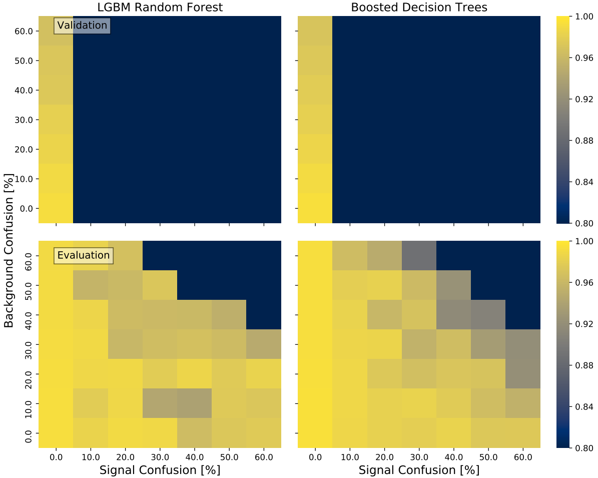
Label Confusion

Label Confusion

This concept was tested by simulating label confusion on a subset of the truthmatched MC photon dataset. The background was chosen to be hadrons, in order to best simulate the background found in data.
The training and validation sets were designed as equal amount of signal and background and exposed to increasing levels of random signal and background confusion (0, 10, 20, 30, 40, 50, 60%), while the evaluation set remained truth labeled.

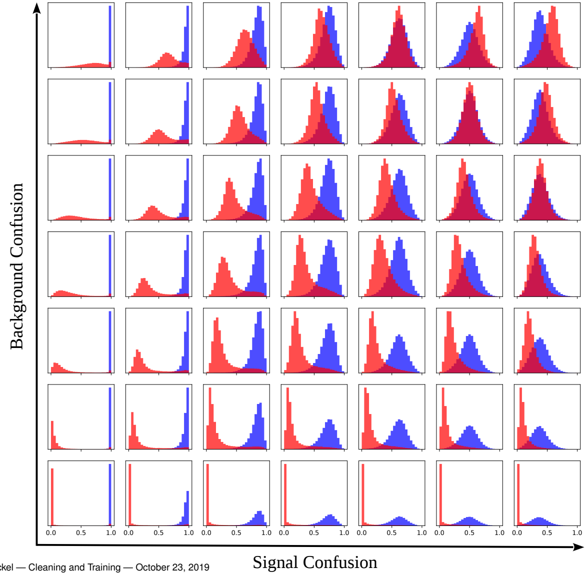
Set	Run 1	Run 2	Run 3
Train	20000	500000	2000000
Validation	10000	50000	100000
Evaluation	157797	1647348	1647348

Run 1 Evaluation

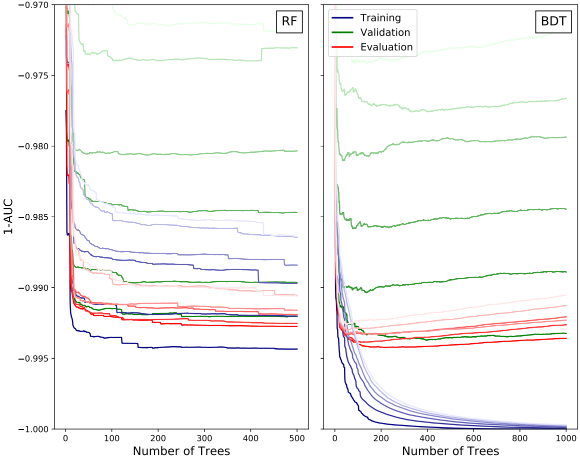


Evolution of metric as a function of increasing confusion, for validation and evaluation sets.

Run 1 Evaluation

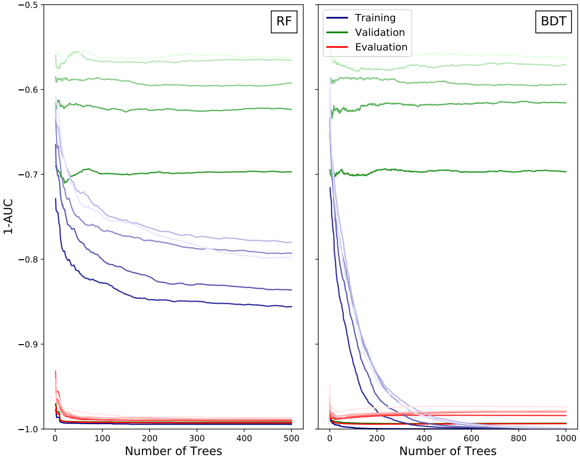


Run 1 Evaluation



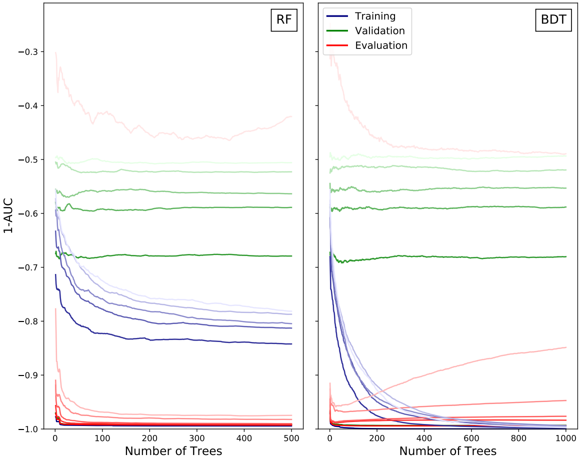
Evolution of metric as a function of building trees with increasing **background** confusion, for the training, validation and evaluation sets.

Run 1 Evaluation



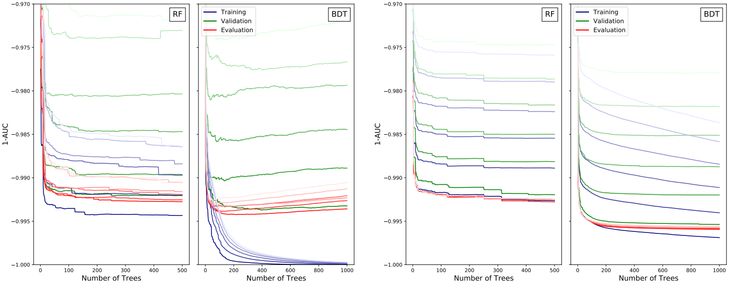
Evolution of metric as a function of building trees with increasing **signal** confusion, for the training, validation and evaluation sets.

Run 1 Evaluation



Evolution of metric as a function of building trees with increasing **symmetric** confusion, for the training, validation and evaluation sets.

Label Confusion

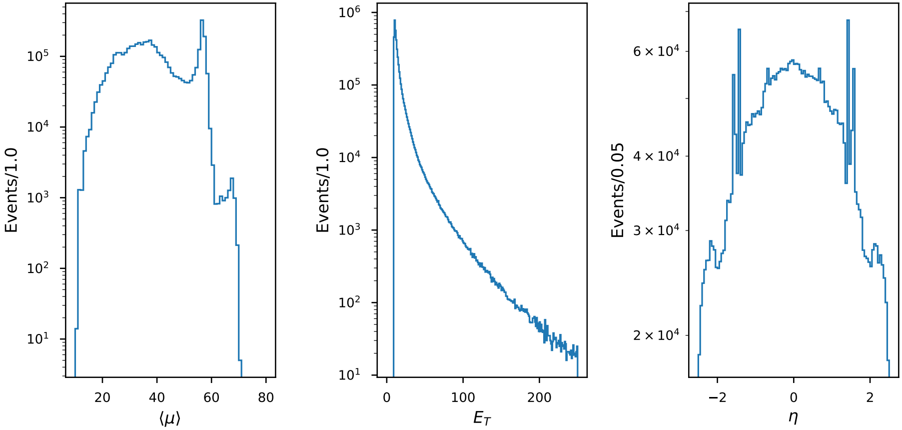


(a) Run 1 (b) Run 3

Evolution of metric as a function of building trees with increasing **background** confusion, for the training, validation and evaluation sets.

DATA Photon dataset Distributions

DATA Photon Dataset



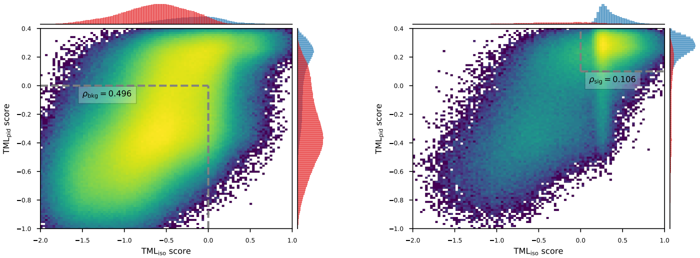
Decorrelation Framework

Decorrelation Framework

The correlations ρ and standard deviations σ is different from the signal and background distributions. This means one has to measure ρ_{bkg}, σ_{bkg} and ρ_{sig}, σ_{sig} on their respective distributions with as little contamination from the other as possible. Signal and background distributions are isolated using the following cuts:

$$Signal : m_{\ell\ell} < 83 GeV \quad and \quad 86 GeV < m_{\ell\ell\gamma} < 96 GeV$$

$$Background : m_{\ell\ell} > 83 GeV \quad and \quad m_{\ell\ell\gamma} > 101 GeV \quad {}^{2}$$



[2] Wrongly quoted in thesis

Decorrelation Framework

The background and signal correlations and standard deviations can be combined in many ways and the following ways have been attempted:

- **Average**:

$$\rho_{pid,iso} = \frac{1}{2}(\rho_{bkg} + \rho_{sig}) \quad \sigma_{pid} = \frac{1}{2}(\sigma_{pid,bkg} + \sigma_{pid,sig}) \quad \sigma_{iso} = \frac{1}{2}(\sigma_{iso,bkg} + \sigma_{iso,sig})$$

- **Piecewise**: Decorrelate using ρ_{sig} and σ_{sig} above and ρ_{bkg} and σ_{bkg} below a threshold in isolation.
- **Continuous**:

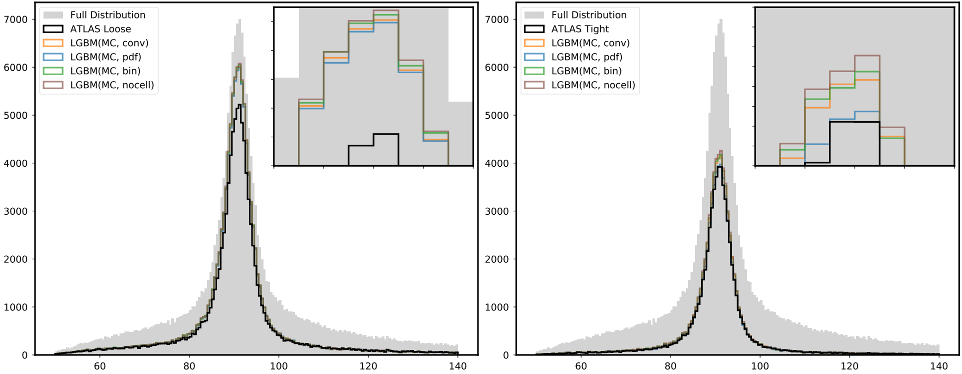
$$f(iso) = \alpha iso + \beta \qquad \alpha = \frac{x_{sig} - x_{bkg}}{\mu_{sig} - \mu_{bkg}} \qquad \beta = x_{sig} - \mu_{sig} \times \alpha$$

- **Sigmoid**:

$$f(iso) = x_{sig} + \frac{x_{bkg} + x_{sig}}{1 + e^{(iso+\mu)/\sigma}} \qquad \mu = \frac{1}{2}(\mu_{sig} - \mu_{bkg}) \qquad \sigma = \frac{1}{2}(\mu_{sig} - \mu)$$

Results in DATA

Results in DATA



The **MC** trained models evaluated on the $Z \to \ell\ell\gamma$ peak, with background efficiency matched to the Loose (left) and Tight (right) working points.

Results in DATA

Improvement [%] (ATLAS Loose)

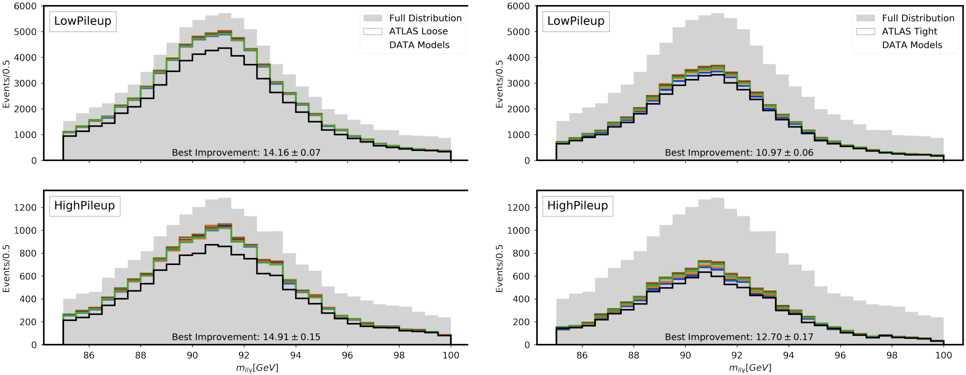
DATA Models	pdf	conv	bin	nocell
initial	12.10 ± 0.05	12.60 ± 0.05	13.94 ± 0.06	14.33 ± 0.06
loose	11.30 ± 0.05	11.82 ± 0.05	12.91 ± 0.06	13.24 ± 0.06
tight	11.23 ± 0.05	11.62 ± 0.05	12.70 ± 0.05	13.14 ± 0.06
MC Models				
Truth	12.59 ± 0.05	12.92 ± 0.06	13.78 ± 0.06	14.08 ± 0.06

Improvement [%] (ATLAS Tight)

DATA Models	pdf	conv	bin	nocell
initial	4.61 ± 0.03	7.45 ± 0.04	9.06 ± 0.05	10.87 ± 0.06
loose	4.92 ± 0.03	7.13 ± 0.04	10.00 ± 0.05	11.17 ± 0.06
tight	4.31 ± 0.02	6.66 ± 0.04	9.25 ± 0.05	10.97 ± 0.06
MC Models				
Truth	1.42 ± 0.01	5.74 ± 0.03	6.47 ± 0.04	7.96 ± 0.04

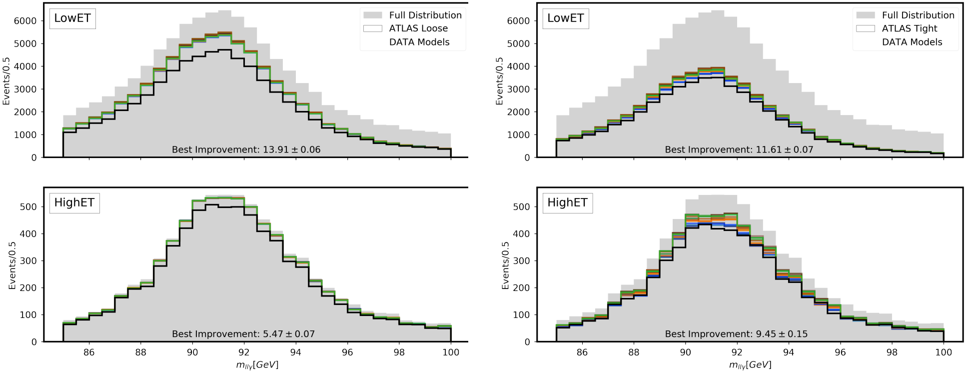
Improvement factor of all models evaluated on the $Z \rightarrow \ell\ell\gamma$ peak, with background efficiency matched to the Loose (top) and

Results in DATA



Split DATA $Z \to \ell\ell\gamma$ evaluation of DATA models at low pileup ($\langle\mu\rangle < 50$) and high pileup ($\langle\mu\rangle > 50$). Compared to the Loose (left) and tight (right) working point.

Results in DATA



Split DATA $Z \rightarrow \ell\ell\gamma$ evaluation of DATA models at low E_T ($E_T < 30GeV$) and high E_T ($E_T > 30GeV$). Compared to the Loose (left) and tight (right) working point.

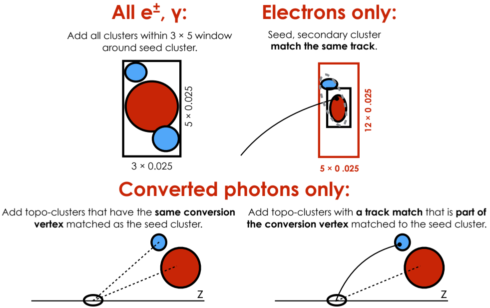
Event Reconstruction

The ATLAS Experiment

Event Reconstruction is done in several steps.

Hits in the Inner Detector are clustered into tracks and deposits in the calorimeter are clustered into clusters.

Photon and Electron Reconstruction has its foundation is these ID tracks and ECAL clusters.



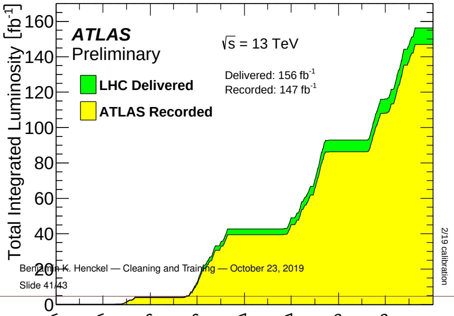
Theory

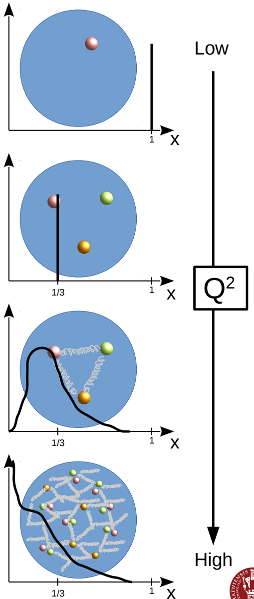
Collider Physics

Natural units: $\hbar = c = 1$. Masses, energies, momenta in eV and lengths and times in eV^{-1}.
$1eV = 1.602 \times 10^{-19} J$.

Coordinate system using the azimuthal angle ϕ and rapidity $y = \frac{1}{2} ln \left[\frac{E + p_z}{E - p_z} \right] \xrightarrow{m=0} \eta = -ln(tan(\theta/2))$

Luminosity is a direct measure of your ability to probe interesting physics





Luminosity

The formula introducing luminosity[3]

$$\frac{d\sigma}{d\Omega} = \frac{1}{I_b \rho dx} \frac{dL\sigma}{d\Omega}$$

where I_b is the current of beam particles, ρ is the density of the target, dx is the target thickness and L is the luminosity. It describes the effective cross section of interaction and is measured in barns ($1 barn = 10^{-24} cm^2$), it has the dimension of area.

$$\frac{N_{events}}{s} = L\sigma \quad , \quad L = \frac{n_{bunches} N_1 N_2 f_{rev}}{A}$$

Where $N_{1(2)}$ is the number of events in the corresponding bunch, A is the overlapping area of the two bunches colliding head on, and f_{rev} is the frequency of revolution.

[3] Hansen, 2015, Particle Detectors and Accelerators Lecture Notes

Production Cross Section

The QCD master formula[4] for production cross sections at LHC, based on the QCD factorization theorem.

$$\sigma = \sum_{a,b=q,g} \int_0^1 dx_a dx_b f_a(x_a, \mu^2) f_b(x_b, \mu^2)$$

$$\times \frac{1}{2\hat{s}} \int_{cuts} \prod_{i=1,n} \frac{d^3 p_i}{2E_i (2\pi)^3} (2\pi)^4 \delta^{(4)}(p_a + p_b - \sum_i p_i) \bar{\sum} |\bar{M}^{ab\rightarrow 1,\ldots,n}|^2$$

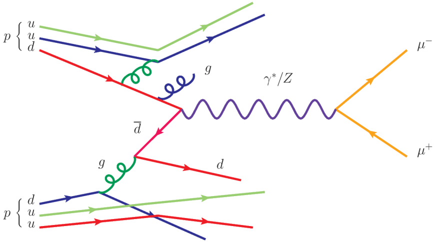


Figure from: https://www.quantumdiaries.org/2015/05/18/dy-resummation/

[4] Thomson, 2013, Modern Particle Physics